DOCUMENT RESUME

ED 253 954 EA 017 535

AUTHOR Smith, Jana K.
TITLE Cost-Outcome Analysis: Measuring Outcomes. Evaluation Guides: Guide Number 4.
INSTITUTION Northwest Regional Educational Lab., Portland, OR. Research on Evaluation Program.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE 84
CONTRACT 400-80-0105
NOTE 31p.; For a related evaluation guide, see EA 017 534.
PUB TYPE Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cost Effectiveness; Educational Assessment; Elementary Secondary Education; *Evaluation Methods; Followup Studies; Input Output Analysis; *Program Effectiveness; *Program Evaluation; Program Validation; *Summative Evaluation

ABSTRACT

The second of two related pamphlets, this guide describes how to design an outcome study and outlines procedures for collecting outcome data. After an introduction that delineates the purpose of the two-part series, the pamphlet is divided into four major sections. The first section explains how to select a cost-outcome analysis. The next three sections address the measurement of utility, benefit, and effectiveness outcome data, respectively. Each self-contained section describes (1) the outcome measure, (2) when to use the analysis, (3) strengths and limitations of the analysis, (4) an example of the analysis, (5) an assessment of the reliability and validity of the outcome data, (6) how to calculate cost-outcome ratios, and (7) how to interpret the cost-outcome ratios. A list of references is included. (TE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

EA

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED253954

EVALUATION GUIDES

Guide Number 4

COST-OUTCOME ANALYSIS: MEASURING OUTCOMES

Jana K. Smith

The main topics discussed include:

- Purpose of This Cost-Outcome Guide Series
- Selecting the Appropriate Cost-Outcome Analysis
- Cost-Utility Analysis
- Cost-Benefit Analysis
- Cost-Effectiveness Analysis

Research on Evaluation Program Nick L. Smith, Director

Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue • Portland, Oregon 97204 • Telephone (503) 248-6800

EA 017 535

1984 2

PURPOSE OF THIS COST-OUTCOME GUIDE SERIES

This guide series was developed in response to evaluators' reported need for information on how to conduct cost-outcome analyses. A recent study on the use of cost-outcome analyses by state education agency (SEA) evaluators (Smith, N. L. & Smith, J. K., 1984) and local education agency (LEA) evaluators (Smith, J. K. 1984) showed

1. in the next five years, 60 percent of all SEAs expect to be asked to conduct cost-outcome analyses,

2. 71 percent of metropolitan LEAs expect to be asked to conduct cost-outcome analyses, and

3. one of the primary impediments to the conduct of cost-outcome analyses is the absence of useful guides and resources.

To assist these and other evaluators in conducting cost-outcome analyses, two "how-to" guides were developed. Although several texts currently exist on cost-outcome anaysis, they tend to be written in technical language or do not give sufficient attention to collection of outcomes as well as costs. This guide series supplements the existing texts by providing concise, readable explanations on how to conceptualize and conduct cost-outcome studies for program evaluation.

The first guide introduced four types of cost-outcome analysis, directed the collection of cost data, and explained how to select the most appropriate cost-outcome analysis to answer an evaluation question.

This second guide describes how to design an outcome study and outlines the procedures for collecting outcome data. In this guide, specific attention is paid to the collection of utility, benefit, and effectiveness data, and to the interpretation of study results.

When Should This Cost-Outcome Guide Series Be Used?

Cost-outcome analyses can be applied in a variety of contexts to answer many different evaluation questions. The procedures are appropriate for evaluating mental health, health, and education programs located in community mental health centers, hospitals, schools, businesses, or any type of service program.

Because of the wide range of possible applications, it was necessary to focus these guides on certain applications. Specifically, this guide was written to direct those evaluation studies that

- are conducted at the local or state level,
- use experimental or quasi-experimental methods,
- are conducted to compare two or more programs,
- measure program outcomes.

The guides are intended to assist the state or local level evaluator conduct a cost-outcome analysis for the purpose of programmatic improvement. For example, those interested in conducting a large-scale national study should consult the reference list of this guide for more technical information (see Levin, 1983; Sugden & Williams, 1978; Thompson, 1980).

Introduction to This Guide

This guide contains four major sections. The first section explains how to select a cost-outcome analysis. The next three sections address the measurement of utility, benefit, and effectiveness outcome data. The sections are self-contained, and can be read separately from the other sections. Each describes

- the outcome measure,
- when to use the analysis,
- strengths and limitations of the analysis,
- an example of the analysis,
- an assessment of the reliability and validity of the outcome data,
- how to calculate cost-outcome ratios and,
- how to interpret the cost-outcome ratios.

After reading the section on selecting an outcome measure, turn to the section pertaining to the outcome measure that best answers your evaluation question.

SELECTING THE APPROPRIATE COST-OUTCOME ANALYSIS

The first steps of a cost-outcome study are identical to the first steps taken in any evaluation study. For example, the evaluation problem must be clarified with program administrators before any design can be suggested. Before a cost-outcome analysis is selected, the evaluator must

- clarify the evaluation question(s),
- identify program goals and objectives relative to the evaluation question(s),
- select outcome variables to be measured, and
- determine whether the measures can be obtained.

These steps are a part of any evaluation, and therefore are undoubtedly second nature to many readers.

To clarify the evaluation question(s), the purpose of the study and the intended use of the results must be ascertained. This information can be obtained through informal conversations with the decision makers. The next step, identifying the program goals and objectives, is actually one of the most difficult and important parts of an evaluation (Rutman, 1980). Without consensual agreement among decision makers and program staff as to the expected program goals and objectives, the evaluation would be directionless, and the results would probably not be utilized. Patton (1978) suggests using interviews, questionnaires, and even decision analysis to elicit and prioritize program goals and objectives.

Once the goals have been specified, it is up to the evaluator to suggest methods to best measure the objectives. Consider this simplistic example:

Question: Which computer programming course should we retain--the lab class or the lecture class?

GOAL FOR LAB CLASS: to teach computer programming

GOAL FOR LECTURE CLASS: to teach computer programming

PRIMARY OBJECTIVE FOR BOTH CLASSES: can write three utility programs

MEASURE OPTION 1: written classroom test

MEASURE OPTION 2: course completion rate

MEASURE OPTION 3: teacher rating

MEASURE OPTION 4: student ratings

Now the evaluator must select the measure that is most reliable, valid, and available. For instance, althougn option 3 may be most valid, it may be the least available measure.

How Do I Select a Cost-Outcome Analysis Method?

The selection of a cost-outcome analysis is contingent upon the purpose of the study and the outcome measure. The difference between the three cost-outcome methods is the outcome measure used in each analysis. In brief, cost utility analysis is based on outcome measures which are estimated (e.g., estimate of the probability of achieving a certain level of self-esteem). Cost-benefit analysis is based on outcome measures which are valued monetarily (e.g., income). Cost-effectiveness analysis is based on outcomes which are measured in effectiveness units (e.g., test scores).

Given an outcome measure, it is simple to find the appropriate analysis. Figure 1 shows an example of the steps undertaken in selecting a cost-outcome analysis for comparing two hypothetical computer programming courses in a prison. The two computer programming courses share the same goals and objectives, so only one flowchart is provided.

To select the cost-outcome analysis most appropriate for a given study, match to the analysis type the outcome measure which best answers the evaluation question. If you are evaluating two programs with similar outcomes, this step is simple. For example, if the question was "Which course is better at reducing recidivism in prisoners?" a cost-effectiveness analysis would be most appropriate. If the question was "Which course is better at increasing the standard of living of prisoners once they are released?" then a cost-benefit analysis would be most appropriate.

This logic also applies to the evaluation of two different programs which share a common outcome. For instance, suppose a counseling program and computer training program are both in effect in a prison and the officials want to know which program is better at reducing recidivism. Although the programs differ, an

effectiveness measure for both could be a count of recidivism for participants. Here a cost-effectiveness analysis could be used, or a cost-benefit analysis, where benefit could be savings accruing from non-recidivism. If you are evaluating programs with different goals, an additional step is necessary to identify the most appropriate cost-outcome analysis.

Figure 1
Selecting a Cost-Outcome Analysis to Compare
Two Computer Programs Which Share the
Same Goals and Objectives

Goals	Objectives	Unit of Outcome Measure	Possible Outcome Measures	Appropriate Cost-Outcome Analysis
to teach computer programming	can write payroll program	effectiveness	test	CE
	can write miscellaneous business utility programs	effectiveness	test	CE
	job placement	effectiveness	count	CE
to rehabilitate inmates	non-recidivism	effectiveness	count	CE
	increase relf-esteem	utility	expurt estimate	CU
	increase standard of living	benefit	increase in income	CB

What If I Am Evaluating Programs With Different Outcomes?

The logic shown in the flowchart in Figure 1 should be followed for each program being evaluated, since programs with different outcomes may be better suited for different types of cost-outcome analysis. Fortunately, selection of a cost-outcome analysis given different program outcomes is not problematic, since cost-benefit and cost-utility methods allow comparisons between programs with different outcomes. (Because these two methods convert outcomes to a common unit, that is, financial benefits or utility estimates, different outcome measures are made comparable.) Nagel (1983) provides an interesting analysis of uses of cost-benefit and cost-utility methods. Only cost-effectiveness analysis requires that outcome measures be alike for all programs being compared.

In general, there are two situations in which the evaluator must select a utility or a benefit analysis because program outcomes are different. These two situations are shown in rows one and two of Figure 2. The first row in Figure 2 shows a situation where the outcome for one program is valued monetarily and the outcome for another is valued in utility estimates. In this case, the evaluator can conduct either a benefit analysis or a utility analysis (row one of Figure 2). For example, suppose the outcome for program A is ideally a monetary measure (e.g., income from a vocational educational program) and the outcome for program B is ideally a utility measure (e.g., improved health as a result of a dental prevention program).

One way to resolve this dilemma is that income can be translated into a utility measure (e.g., the value of increasing the standard of living 10 percent) and compared to the utility value of improved health. Or another way is to translate improved health into a benefit measure (e.g., amount of money saved on dental care) and compared to the increase in income resulting from the vocational program. Which outcome is converted depends upon the evaluation question and the analysis which would be most meaningful to the decision maker.

The second row in Figure 2 shows a situation where the outcome for one program is valued monetarily and the outcome for another is valued in terms of effectiveness. Either the effectiveness measure can be converted into a monetary benefit, if possible, or both the monetary benefit and effectiveness outcomes can be converted into utility values. This conversion could be accomplished by estimating the utility of the benefit measure and comparing it to the estimated utility of the effectiveness measure.

Rows three and four of Figure 2 demonstrate the situations where utility analysis is the only possible analysis to answer the evaluation question when outcomes for two programs are different. In row three, effectiveness data is only available for Program A. That is, while the outcomes in program B could be measured in terms of effectiveness, the data are not available. Instead, the outcomes in program B are estimated in utility. Therefore, to make outcomes comparable, program A outcomes must be estimated as utilities also.

In row four, the effectiveness measures for two programs are entirely different but comparisons are requested by the decision maker (e.g., compare the effectiveness of the reading program and math program). In this situation, outcomes must be converted into subjective utility values to be made comparable. For example, math and reading scores are not comparable, but the value of increasing math scores 10 percent can be compared to the value of increasing reading scores 10 percent.

Figure 2
Selecting a Cost Outcome Analysis for
Programs with Different Outcomes

Outcome		Analysis
Program A	Program B	
monetary	utility	benefit or utility
monetary	effectiveness	benefit or utility
effectiveness	utility	utility
effectiveness (e.g., reading)	effectiveness (e.g., math)	utility

Determining Whether Reliable and Valid Outcome Measures Can Be Obtained

The design of a reliable and valid outcome evaluation is extremely complex, and many books have been written to improve the conduct of evaluation studies[1]. Obviously, the intent of this guide is not to compete with such thorough works on the conduct of evaluation research. The reader is encouraged to consult any of these books for guidance in the design of an evaluation. The purpose of this guide is to assist the evaluator in obtaining meaningful outcome measures to use in a cost-outcome analysis.

The next step is to design and conduct the evaluation. The remainder of this guide describes the problems of designing and conducting a cost-utility, cost-benefit, and cost-effectiveness analysis.

COST-UTILITY ANALYSIS

Cost-utility (CU) analysis compares the costs of programs to outcomes which have been estimated. Utility, or the subjective value of an outcome or outcomes, can be generated by experts, program sponsors, or program participants for any type of outcome that is meaningful to the decision maker. The utility measure can

---

[1]See Berk, 1981; Conner, 1981; Cook & Campbell, 1979; Datta & Perloff, 1979; House, 1980; Kalton, 1983; Patton, 1978, 1981, 1982; Rutman, 1977, 1980; Schulberg & Jerrell, 1979; Smith, 1981, 1981, 1982; and Weiss, 1972, 1975.

then be combined with cost data to form a cost-utility ratio. A rank ordering of cost-utility ratios for two or more programs will indicate which program has more utility relative to its costs.

Description of Utility Measures

Utility measures

- can be generated for any type of program outcome,
- can assess unidimensional or multidimensional outcomes for a given program, and
- can be used to compare programs with different outcomes.

Common measures might include the perceived utility of

- reading at the 8th grade level,
- completing a program,
- increasing self-esteem, or
- being placed in a job.

Measures of utility can be unidimensional or multidimensional. An example of a unidimensional measure is parents' rank order of their preferences for a group of different school-sponsored programs (Levin, 1983). These rankings are averaged and combined with cost data to form cost-utility ratios which, when ranked, show which programs have the most cost-utility and which programs have the least. These comparisons of programs with different goals are possible because they were all valued in the common unit of utility.

An example of a multidimensional measure is the rating of two alternative reading programs on four dimensions: increasing reading speed, increasing reading comprehension, increasing word knowledge, and increasing student satisfaction with reading. For each dimension, respondents rate their perceived utility on a scale of 0 (least important) to 10 (most important). These ratings are averaged, and the cost utility of the two alternatives are compared, using composite utility measures.

When to Use Cost-Utility Analysis

Cost utility is a valuable planning and decision-making tool, particularly when

- outcome data are not available,
- time or money limitations restrict collection of outcome data,

- immediate assessment of outcome is needed to facilitate discussion,

- it is desirable to take into account the opinion of experts or program participants.

The Limitations of Cost Utility Analysis

While cost utility is viewed as a useful analysis technique, proponents of the techniques are quick to point out the problems associated with using subjective estimates rather than measured outcome data for program evaluation and decision making. For instance, because utilities are generated by those invested in a particular program, the results would be meaningless to other programs.

There have also been questions about the meaningfulness of the utility measures for internal decision making. Utility is a measure of human values; consequently, estimates of program outcomes are more likely to reflect personal biases than actual program functioning. Such subjective ratings have been shown to be unreliable and invalid. Following are examples.

- Personal preferences are often "intransitive" (Tversky, 1969) in that preference for A over B and B over C does not insure that A is preferred over C. Consequently, it is often not possible to determine individual preference based on comparative ratings.

- People differ in the way that they scale their preferences. For example, on a scale of 1 to 7, with 7 being most preferred, Joe may rate his most preferred programs as 6 and 7. Sally may rate her most preferred programs as 4 and 5. Obviously, a comparison of such ratings could be misleading.

- Finally, personal biases, conscious distortions, or caprice may affect the validity of the ratings.

Collecting the Utility Data

Texts on cost-outcome analysis have been negligent in providing a detailed description of how one actually goes about collecting utility data. The best available source for such direction can actually be found in the literature on decision analysis. While decision analysis is a widely recognized method of deriving utility measures for decision making, cost analysts have been slow to take advantage of the methods for use in cost-utility analysis. The utility methods, developed by decision

analysts, have been specifically designed to overcome many of the problems with reliability and validity associated with subjective ratings mentioned previously. Indeed, use of decision analysis methods can generate utility measures which are meaningful and useful (Edwards, Guttentag & Snapper, 1975; Pitz & McKillip, 1984).

Using Decision Analytic Techniques to Generate Utility Measures

Two decision analysis methods have been suggested as appropriate for evaluation of outcomes. These are MAUT, multiattribute utility theory, (Edwards, Guttentag & Snapper, 1975) and SMART, simple multiattribute rating technique (Edwards, 1977). The techniques used in these two methods are very similar, and both were developed specifically to accommodate the needs of busy decision makers working under short timelines. The steps involved in carrying out these methods are listed below. Detail about each step is provided in the example in the next section of this guide.

Checklist of Steps in the Decision Analytic Generation of Utilities

Step 1: Identify the Decision Maker

Step 2: Identify the Decision Question

Step 3: Identify the Program Goals

Step 4: Identify Relevant Program Outcomes to be Estimated

Step 5: For Each Program, Rank Order the Outcomes in Order of Importance

Step 6: Rate the Outcomes to Obtain Importance Weights

Step 7: Normalize the Importance Weights: Divide Each by the Total Sum of Importance Weights

Step 8: Estimate the Probability of the Occurrence of Each Outcome

Step 9: Calculate the Utility--Importance X Probability for Each Goal

Step 10: Make a Decision Recommendation

An Example of a Cost-Utility Analysis

Because utility measures are subjective and program specific, cost-utility analyses are not published, and therefore no known examples are available. There are a few examples of MAUT analyses (e.g., see Edwards, et. al, 1975; Edwards, 1977; Pitz and McKillip, 1984). There are, however, no simple examples of MAUT for program evaluation purposes; consequently, the following hypothetical example was developed.

Because of unexpected, but not surprising, financial cutbacks, the County School Board must decide whether to discontinue the Driver's Education Program or the Dance Theater Program. Driver's Education has been offered for 10 years, and is required by the state in order to receive a driver's license at age 16 rather than age 18. The Dance Theater Program has only been in effect for three years, and has been extremely popular among the school's minority students. For the sake of simplicity, assume that both programs provide service for 100 students a term.

The first three steps of the analysis were straightforward. The remaining steps, however, required the aid of a group of students, parents, and faculty. All of the actions resulting from completion of steps four to nine are depicted in Table 1.

Step 1: Identify the decision makers.
The decision makers are the school board.

Step 2: Identify the decision question.
Which program should we discontinue?

Step 3: Identify the program goals.
The program goals were obviously different. Driver's Education attempts to teach driving skills, while Dance attempts to teach dance skills.

Step 4: Identify relevant program goals.
Students, parents, and faculty were called in to help identify which goals of the programs should be considered. Several unexpected goals were identified for each program, as shown in column 1 of Table 1.

Step 5: Rank the goals for each program.
Again, students, parents, and faculty helped to order the goals by importance. Group consensus methods were used in arriving at the final order, although individual rankings could have been derived and averaged. The rank order of outcomes is shown in column 2 of Table 1.

Table 1

Steps 4 to 9 of a Utility Analysis

Program	Step 4 Identify Goals	Step 5 Rank Goals	Step 6 Rate Goals	Step 7 Calculate Importance Weight	Step 8 Judge Probability	Step 9 Utility (Importance X Probability)
Driver's Education	Reduce traffic fatalities	Pass state exam	10	10/150 = .06	.90	.05
	Reduce drinking & driving	Reduce drinking & driving	50	50/150 = .33	.40	.13
	Pass state driver's exam	Reduce fatalities	90	90/150 = .60	.70	.42
			150			.60
Dance Theater	Increase self-esteem	Teach dancing	10	10/80 = .12	.40	.05
	Reduce vandalism	Reduce vandalism	20	20/80 = .25	.20	.05
	Teach dance	Increase self-esteem	50	50/80 = .62	.50	.31
			80			.41

Step 6: Rate the goals according to importance.

To achieve the closest possible approximation to ratio level ratings, raters (again, students, parents, and faculty) were asked to assign their least preferred outcome a value of 10. Each subsequent rating was then to be judged in terms of "How many more times important is this outcome than the least preferred outcome?" That is, (refer to Table 1) if passing the state driving exam is rated a 10 . . . how many more times important is the reduction of drunken driving? In this case, after group discussion, it was decided that reduction of drunken driving was five times more important than passing the state driving exam.

Step 7: Sum the importance weights and divide each by the total importance sum

This step is shown in the fourth column in Table 1.

Step 8: Estimate the probability of the occurence of each outcome.

These probabilities were again group generated, although the faculty involved in each program provided most of the information about historical program effectiveness. Probabilities were generated as percent chance of being accomplished. For example, the group agreed that while most students eventually pass the state exam (90 percent chance of success), far fewer actually refrain from drinking and driving after taking the program (40 percent chance of reducing drinking and driving). See the fifth column for probabilities generated in this exercise.

Step 9: Calculate the utility.

The utility of each outcome was calculated by multiplying the importance weights by the probabilities as shown in the last column of Table 1. These utilities were then summed to form a composite utility for each program.

Step 10: Make the decision.

In utility analysis, the larger the utility, the greater the perceived value of the program. In our example, the Driver's Education Program had a utility of .60, as compared to a utility of .41 for the Dance Theater Program. These figures, however, should not be interpreted for decision making purposes until the reliability and validity of the data have been considered. The next two questions deal with these important issues.

How MAUT Improves the Integrity of the Utility Data

Recall the three criticisms of utility analysis having to do with rating credibility: intransitivity, scaling differences, and rater caprice. Systematic application of MAUT helps to minimize these problems. The use of group input should minimize the chance of intransitivities, since group interaction is more likely to catch such an inconsistency than would an individual rater. The calculation of the weighted importance values eliminates the problem of scaling differences. Finally, group interaction would also help insure minimal caprice in the ratings. In fact, a group decision, or individual ratings averaged, would help insure stability of ratings.

Assessing the Reliability and Validity of the Utility Data

The validity of the data is contingent upon the generation of meaningful program outcomes, while the reliability concerns consistency of the numerical data associated with those outcomes. Levin (1983) suggests that, to enhance validity, it is important that the utility ratings be generated by those who are most familiar with the program, and who will be affected by the changes. In our example, obtaining information from students, parents, and faculty would help to insure that the utilities are a valid reflection of the values of individuals involved in the program. Similarly, using group processes (or averaging individual data) would increase the chance that the ratings are stable (reliable). Individual ratings are much more likely to be unrealizable than are group ratings.

Calculating the Cost-Utility Ratios

Cost-utility ratios are calculated by combining the cost data with the utility data. Suppose, in this example, the total program cost was calculated at $2,000 for Driver's Education and $1,000 for Dance Theater. As calculated in Table 1, the utility for Driver's Education is .60 and the utility for Dance Theater is .41. (In order to obtain more intuitively meaningful figures, these utilities were multiplied by 100 for this example.) The cost utility of Driver's Education is 2000/60 = 33.33, and for Dance Theater is 1000/41 = 24.39. These figures are shown in Table 2. Notice these are program level data: total cost versus total utility. Per-participant level data can also be generated by dividing both total figures by the number of participants It is important not to mix total figures with per-participant figures. For utility analysis, however, program level data is often most meaningful.

Table 2
Calculating the Cost-Utility Ratio

Program	Total Program Cost	Composite Utility Measures X 100*	CU Ratio
Driver's Education n = 100	$2,000	.60 X 100 = 60	33.33
Dance Theater n = 100	$1,000	.41 X 100 = 41	24.39

*Multiplied by 100 to obtain whole numbers.

Interpreting the Cost-Utility Ratios

The cost-utility (CU) ratio is useful only when compared to another CU ratio. There is no intuitive interpretation of a singular ratio. Recall that a higher utility is preferred when just examining utility values. However, when comparing two ratios, the lower ratio reflects less cost per unit of utility. In this case, while the utilities for Driver's Education is higher than for Dance Theater, the CU analysis points to the retention of the Dance Theater rather than the Driver's Education Program.

COST-BENEFIT ANALYSIS

Cost-benefit (CB) analysis compares the costs of programs to outcomes which have been valued monetarily. This method has often been used in the assessment of nationwide health policy, and in environmental policy studies (Thompson, 1980). It has had fewer applications in program evaluation. One reason that cost-benefit analysis has not been widely used in program evaluation is that it is difficult to assign monetary values to social service program outcomes that are not traded in the market (Thompson, 1980). For example, it is difficult to monetarily value an increase in reading comprehension.

Description of Benefit Measures

Thompson (1980) has identified five categories of methods for valuing program outcomes monetarily. These categories are as follows.

1. Money measurement.

This is a simple assessment of the amount of money saved or earned as a result of participating in a program. It is the simplest and most straightforward valuing method. Examples of this type of measure include:

Evaluation Question	Outcome Measure
How much does participation in vocational education increase the student's income?	income
How much does the health screening program save participants in terms of income lost due to illness?	income saved
How much money is saved by taking a bus versus a commuter train?	money saved

2. Market valuing.

This is very similar to money measurement, except that it involves the transformation of an outcome into a monetary value based on market rates. For example, the cost of a new school might be compared to the benefit of increasing land value on the adjacent real estate.

3. Econometric estimations.

This valuing method requires an understanding of the supply and demand curve, and is not easily calculated by evaluators. Econometric estimations are most often used to figure the value of a human life for federal level policy decisions (Rhoades, 1978). In brief, there are several ways to estimate how much a life is worth, such as

- own life value — how much an individual will pay to live
- acquaintanceship life value — how much an acquaintance would pay to enable a person to live
- pareto life value — the maximum that could be paid to keep someone alive without negatively affecting someone else
- consumptive life value — life value depends upon how much money is spent
- earnings life value — how much an individual would earn in a lifetime

- willingness to pay — rational average someone will pay for chance to live
- social life value — amount a group of people think the government should pay to keep someone in treatment

4. Hypothetical questions.

   This method asks an individual or group a hypothetical question, such as "How willing would you be to pay for this service . . . or to avoid this tragedy . . . or to attend a session on this versus that topic?" For example, a willingness-to-pay question is, "How much would you be willing to pay to attend a session on assertiveness training versus a session on hypnosis?"

5. Observing political choices.

   This method uses social and legislative action as a gauge for the financial value. For example, if the equal rights amendment passes, then the same methods for valuing a life based on life earnings can be used for men and women. Until then, however, different methods must be used to calculate earnings for men and women.

There are significant problems with all of these methods, and their use is generally not recommended (Thompson, 1980). Except for money measurement and market valuing, most methods are problematic in that they contain systematic biases, underestimations, and distortions. Rothenberg (1975) commented eloquently on the state of data for cost-benefit analysis:

> . . . regardless of its [cost-benefit analysis] methodological claims, its practical usefulness will be most decisively at the mercy of the availability of data. Very serious inadequacy of relevant data exists in almost every area for which cost-benefit analyses have been undertaken. (p. 88)

When to Use Cost-Benefit Analysis

There are times when program outcomes are verbalized in terms of monetary value, such as in the evaluation of a vocational education program, to see its effect on the incomes of participants. Cost-benefit analysis is most useful for evaluating programs with outcomes which can logically be valued monetarily, or when the decision makers specifically request financial outcome data. In addition, because outcomes are valued in a common unit (monetary value), programs with different outcomes are

comparable. So, like cost-utility analysis, a mechanic's program could be compared to a program for unwed high school mothers if the monetary value of each program can be meaningfully determined.

Example of Cost-Benefit Analysis

There are not many published examples of cost-benefit analysis conducted for the purpose of program decision making. Most examples are of a policy nature, and assess benefit on a societal rather than programmatic level. Andrieu (1977) provides an interesting example of a cost-benefit analysis of manpower training programs, and uses yearly earnings as a benefit measure. Malitz (1984) recently conducted a cost-benefit analysis of family planning services in Texas. Malitz used econometric estimates of the costs of the number of pregnancies, births, abortions, and miscarriages averted.

As an example of cost-benefit analysis for programmatic decision making, consider this hypothetical decision problem about Driver's Education versus the Dance Theater.

Because of unexpected, but not surprising, financial cutbacks, the County School Board must decide whether to discontinue the Driver's Education Program, or the Dance Theater Program. Driver's Education has been offered for 10 years, and is required by the state in order to receive a driver's license at age 16 rather than age 18. The Dance Theater Program has only been in effect for three years, and has been extremely popular among the school's minority students. For the sake of simplicity, assume that both programs provide service for 100 students a term.

The selection of a benefit measure should be based on a measure that would be meaningful to decision makers. The decision makers have asked which program generates the most revenue for the school. The benefit generated by the Dance Theater can be easily measured by tabulating door charges from shows.

The benefit of the Driver's Education program is not as easily measured. The decision makers solved the measurement problem by wondering if students who could drive would be more likely to attend school events. Since, without driver education, a student cannot take the state licensing exam until the age of 18, it is theoretically possible that by taking driver's education, the student may gain two years of opportunity to attend school events. Benefit for this evaluation could be defined as fees collected from students attending events as a result of having a driver's license. Students who took driver's education last year were surveyed and asked a hypothetical question about their

attendance now versus their estimated attendance if they did not have their license. Increases in school revenue were projected, based on these responses, and fees collected at the school events. The benefit derived from the Driver's Education program was calculated at $600 and for the Dance Theater at $1000.

Evaluating the Reliability and Validity of the Benefit Measures

One of the first issues to consider about the validity of benefit measures is whether they accurately reflect program functioning. Obviously, the direct market value of fees collected at Dance Theater shows is not likely to be questioned. The estimation of events attended as a result of having a driver's license is much more suspect. Prior to the collection of this particular measure, the evaluator should have confirmed with the decision makers that this measure would be meaningful to them. If it is meaningful, then the evaluator should examine the method used for sampling students, and consider the possibility of systematic biases which may affect responses.

Calculating the Cost-Benefit Ratio

Once the benefit and cost data have been collected, it is simple to combine the two to form a cost-benefit ratio. Table 3 demonstrates that the total program cost for Driver's Education ($2,000) when divided by the total benefit ($600) results in a ratio of 3.33. The cost-benefit ratio for the Dance Theater is .33 (1000 ÷ 3000). The interpretation of these figures is in the next section. This example also shows that when student fees are used as a measure of benefit, the Driver's Education program results in a loss of $1,400 (net benefit = benefit - cost), while the Dance Theater earns $2,000.

Notice that these ratios are based on program level data. Total program cost is being compared to total program benefit. It can also be useful to calculate the cost-per-participant for each program. To calculate per-participant figures, divide the total cost by the number of participants.

Interpreting the Cost-Benefit Ratio

The net benefits show that the costs of driver's education far exceed the benefits--when benefits are based on money generated from students' fees. In contrast, the Dance Theater benefits far exceed the costs--when benefits are based on money generated from student fees. Another way to interpret these results is to translate the ratios into cost terms. Driver's Education spent $3.33 to generate $1.00 of income for the school, whereas the

Dance Theater spent only 33¢ to generate $1.00 of revenue for the school. Clearly, if decision makers are primarily concerned with the revenue produced by these programs, they would retain the Dance theacer and discontinue Driver's Education.

Table 3
Calculating a Total Cost-Benefit Ratio

Program	Total Program Cost	Total Benefit	CB Ratio	Net Benefit
Driver's Education n = 100	$2,000	$600	3.33	-1400
Dance Theater n = 100	$1,000	$3,000	.33	+2000

COST-EFFECTIVENESS ANALYSIS

Cost-effectiveness analysis compares the costs of a program to outcomes measured in units of effectiveness. When evaluators think of program outcome measures, they automatically think of effectiveness measures. Historically, effectiveness measures such as test scores, performance measures, or other ratings, have been used as a barometer of how well a program is achieving its goals. Analysis of effectiveness data can be used formatively to improve program functioning, or summatively to make programmatic decisions. Outcomes measured in effectiveness terms are:

- empirical and therefore replicable,
- meaningful to decision makers and participants alike,
- often suitable for complex statistical analyses,
- sometimes available as a result of routine program monitoring.

Descriptions of Effectiveness Outcome Measures

Measures of effectiveness can be obtained for outcomes of cognitive skills and psychomotor skills, and for affective variables, using tests, observation, surveys, and program reccrds. Examples of outcome measures include the following.

- difference between pretest and posttest scores,
- final test scores,
- program completions,
- job placement,
- satisfaction,
- job performance,
- morbidity or mortality,
- number of deaths averted,
- recidivism rate,
- number of clients served.

Obtaining Meaningful Outcome Measures

Any text on evaluation addresses the critical issues of measuring and collecting effectiveness data. The true difficulty in conducting a valid and reliable cost-effectiveness analysis is obtaining meaningful measures. When considering a test, it is best to use one that has been standardized and shown to be reliable and valid (Nunnally, 1975; Nunnally & Wilson, 1975), rather than to develop your own measure. If you do develop a test, there are important procedures necessary to insure that the collected data accurately reflect program effectiveness. Cole and Nitko (1981) outline these important issues that go beyond the scope of this guide. Similarly, for using archival data, interviews, or surveys, Weiss (1975) and Weinstein (1975) have helpful suggestions. Perhaps the simplest effectiveness measures would be tabulations of number of clients served, number of dropouts, or other similar indicators of program effects.

To review, when deciding upon effectiveness measures, consider these questions.

- Is the measure meaningful to decision makers?
- Is the measure available?
- Is the measure reliable?
- Is the measure valid?

Limitations of Cost-Effectiveness Analysis

Cost-effectiveness analysis, unlike cost-utility and cost-benefit analyses, requires that comparisons be made on the same outcomes. For example, it is meaningful to report that at the end of the term students in computer assisted instruction (CAI) reading got 80 percent correct on the final exam, while students receiving peer tutoring in reading got only 60 percent correct on the same exam. Obviously, unless some other variable was affecting student scores, CAI was more effective in increasing reading ability. Such a comparison, however, cannot be made between reading and health programs where students in reading get 80 percent correct on the reading exam and students in health

education get 50 percent correct on the health exam. Since the outcome measures of the two programs are dissimilar, it cannot be concluded that one program is more effective than the other.

Examples of a Cost-Effectiveness Evaluation

There are many examples of CE studies that have been published. A few of these include:

Author	Evaluation Question	Outcome Measure
Doherty & Hicks (1977)	Effectiveness of health-care programs for the elderly	Scale of daily activities
Klarman, Francis & Rosenthal (1968)	Cost-effectiveness of treatment of chronic venal disease	Mortality
Quinn, Van Mondfrans & Worthen (1984)	Cost-effectiveness of two math programs	Achievement test scores
Skellie, Mobley, & Coan (1982)	Cost-effectiveness of long-term health care	Days of survival
Yates, B. (1978)	Cost-effectiveness of weight-loss program	Pounds lost

Reliability and Validity of Effectiveness Outcome Data

If effectiveness outcome data are not reliable (stable) and valid (meaningful measure of program), they will be of little use in evaluation. Cook and Campbell (1979) provide an excellent discussion of reliability and validity issues in quasi-experimental studies. The following is a checklist of issues intended to be used prior to interpretation of data collected in an outcome study where participants are not randomly assigned to programs.

Reliability. Is the outcome measure stable? If outcomes are available for only a few participants, spurious effects may result. However, if data are aggregated for a number of participants, the reliability is increased.

Validity. Internal validity refers to the relationship between the program implementation and the outcome. There are many variables that can complicate the interpretation of an outcome and bring to question whether the outcome measure accurately represents program effectiveness. The following summarizes the discussion about threats to internal validity by Cook and Campbell (1979).

Each threat is described using a hypothetical third-grade reading program which uses the difference between pretest and posttest administration of a standardized achievement test as an outcome measure.

History: Did any other factor contribute to the observed change in the program participant? For example, was a lunch program implemented at the same time that might have affected achievement scores?

Testing: Did students recall the pretest at the time of the posttest and as a result do better?

Instrumentation: Is it possible that the test was too hard ("floor" effect), or too easy ("ceiling" effect) at either the pretest or posttest?

Statistical Regression: If students were selected for participation in the program because of low achievement scores or deprived home environments, then an increase in test scores between the pretest and the posttest could occur simply as a statistical artifact.

Selection: If comparing two similar programs, consider whether the children in the two programs were dissimilar for any reason (e.g., neighborhood differences, age differences, sex differences).

Mortality: Ask if gains might be spuriously distorted because low achieving or high achieving students dropped out of the program.

These are just a few of the basic threats to internal validity. Evaluators unfamiliar with these or other threats to internal validity should consult Cook and Campbell (1979, pp. 53-55) for additional information.

Calculation of Cost-Effectiveness Ratios

Once the outcome data have been collected and examined for reliability and validity, it is a simple matter to combine the cost with the outcome data to form a ratio. Suppose computer assisted instruction and peer tutoring reading programs are being

compared for cost-effectiveness. In this example it cost $500 per student to provide the computer assisted instruction program and the average student got 70 points correct on the final exam. This computes to a cost-effectiveness ratio of 7.1. The peer tutoring program was much less expensive per student ($100) and the average student got 50 points correct on the final exam. This computes to a cost-effectiveness ratio of 2. Table 5 depicts these figures and calculations.

Table 5
Calculating the Cost-Effectiveness Ratio

Program	Program Cost Per Student	Effectiveness Cost Per Student	CE Ratio
CAI reading	$500	70 points correct	7.1
Peer tutoring reading	$100	50 points correct	2

Notice that these figures are represented in per-participant figures. Per-participant figures are determined by dividing program-level data (e.g., costs or outcomes) by the number of participants. It is important not to mix program level costs with participant level outcomes, since the ratio would be misleading. Either per-participant or program level ratios are meaningful; the selection of either should be based on the types of answers most useful to the decision makers.

Interpretation of Cost-Effectiveness Ratios

A cost-effectiveness ratio can be interpreted as the cost to produce a unit of effectiveness. It cost about $7 per point with computer assisted instruction versus only $2 per point with peer tutoring.

Of course, there is a trade-off here. While peer tutoring is less expensive per point, students in computer assisted instruction received an average of 20 more points at the end of the term. This creates an ethical dilemma for the decision makers: now that they have been provided with the data, they must use it wisely in making decisions.

REFERENCES

Andrieu, M. (1977). Benefit-cost evaluation. In L. Rutman (Ed.), Evaluation research methods: A basic guide (pp. 217-232). Beverly Hills, CA: Sage Publications, Inc.

Berk, R. (Ed.) (1981). Educational evaluation methodology: The state of the art. Baltimore: The Johns Hopkins University Press.

Cole, N. & Nitko, A. (1981). Measuring program effects. In R. Berk (Ed.), Educational evaluation methodology: The state of the art (pp. 32-63). Baltimore: The Johns Hopkins University Press.

Conner, R. (Ed.) (1981). Methodological advances in evaluation research. Beverly Hills, CA: Sage Publications, Inc.

Cook, T. & Campbell, D. (1979). Quasi-experimentation design and analysis issues for field settings. Chicago: Rand McNally.

Datta, L. & Perloff, R. (1979). Improving evaluations. Beverly Hills: Sage Publications, Inc.

Doherty, N. & Hicks, B. (1977). Cost-effectiveness analysis and alternative health care programs for the elderly. Health Services Research, 12, 190-203.

Edwards, W. (1977). How to use multiattribute utility measurement for social decision making. IEEE Transactions of Systems, Man, and Cybernetics, Vol. SMC - 7,326 - 340.

Edwards, W., Guttentag, M. & Snapper, K. (1975). Effective evaluation: A decision theoretic approach. In E. Struening & M. Guttentag (Eds.), Handbook of evaluation research Vol. 1. Beverly Hills, CA: Sage Publications, Inc.

House, E. (1980). Evaluating with validity. Beverly Hills, CA: Sage Publications, Inc.

Kalton, G. (1983). Introduction to survey sampling. Beverly Hills, CA: Sage Publications, Inc.

Klarman, H. E., Francis, J. & Rosenthal, G. D. (1968). Cost-effectiveness analysis applied to the treatment of chronic renal disease. Medical Care, 6(1), 44-54.

Levin. H. (1983). Cost-effectiveness: A primer. Beverly Hills, CA: Sage Publications, Inc.

Malitz, D. (1984). The costs and benefits of Title XX and Title XIX family planning services in Texas. Evaluation Review, 8(4), 519-536.

Nagel, S. S. (1983). Nonmonetary variables in benefit-cost evaluation. Evaluation Review, (8)1, 37-64.

Nunnally, J. (1975). The study of change in evaluation research: Principles concerning measurement. In E. Struening & M. Guttentag (Eds.), Handbook of evaluation research, Vol. 1 (pp. 101-138). Beverly Hills, CA: Sage Publications.

Nunnally, J. & Wilson, W. (1975). Method and theory for developing measures in evaluation research. In E. Struening & M. Guttentag (Eds.), Handbook of evaluation research, Vol. 1 (pp. 227-288). Beverly Hills, CA: Sage Publications.

Patton, M. (1978). Utilization focused evaluation. Beverly Hills, CA: Sage Publications, Inc.

Patton, M. (1981). Creative evaluation. Beverly Hills, CA: Sage Publications, Inc.

Patton, M. (1982). Practical evaluation. Beverly Hills, CA: Sage Publications, Inc.

Pitz, G. & McKillip, J. (1984). Decision analysis for program education. Beverly Hills, CA: Sage Publications, Inc.

Quinn, B., Van Mondfrans, A. & Worthen, B. (1984). Cost-effectiveness of two math programs as moderated by pupil SES. Educational Evaluation and Policy Analysis, 6(1), 39-52.

Rhoads, S. E. (1978). How much should we spend to save a life? The Public Interest, 51, 74-92.

Rothenberg, J. (1975). Cost-benefit analysis: A methodological exposition. In M. Guttentag & E. Struening (Eds.), Handbook of Evaluation Research, Vol. 2 (pp. 55-88). Beverly Hills, CA: Sage Publications, Inc.

Rutman, L. (Ed.) (1977). Planning useful evaluations. Beverly Hills, CA: Sage Publications, Inc.

Rutman, L. (Ed.) (1980). Evaluation research methods: A basic guide. Beverly Hills, CA: Sage Publications, Inc.

Schulberg, H. & Jerrell, J. (Eds.) (1979). The evaluator and management. Beverly Hills, CA: Sage Publications, Inc.

Skellie, F. A., Mobley, G. M. & Coan, R. E. (1982). Cost-effectiveness of community-based long-term care: Current findings of Georgia's alternative health services project. American Journal of Public Health, 72(4), 353-358.

Smith, J. K. (1984). Cost analysis at the local level: Applications and attitudes (ROEP Paper and Report Series No. 103). Portland, OR: Northwest Regional Educational Laboratory.

Smith, N. L. (Ed.) (1981). Metaphors for evaluation. Beverly Hills, CA: Sage Publications, Inc.

Smith, N. L. (Ed.) (1981). New techniques for evaluation. Beverly Hills, CA: Sage Publications, Inc.

Smith, N. L. (Ed.) (1982). Communication strategies in evaluation. Beverly Hills, CA: Sage Publications, Inc.

Smith, N. L. & Smith, J. K. (1984). Cost studies in educational evaluation: The state of practice (Research on Evaluation Program Paper and Report Series No. 100). Portland, OR: Northwest Regional Educational Laboratory.

Sugden, R. & Williams, A. (1978). The principles of practical cost-benefit analysis. Oxford: Oxford University Press.

Thompson, M. (1980). Benefit-cost analysis for program evaluation. Beverly Hills, CA: Sage Publications, Inc.

Thompson, M., Rothrock, J., Strain, R. & Palmer, R. (1981). Cost analysis for program evaluation. In R. Conner (Ed.), Methodological advances in evaluation research (pp. 31-46). Beverly Hills, CA: Sage Publications, Inc.

Tversky, A. (1969). Intransitivity of preference. Psychological Review, 76, 31-48.

Weinstein, A. (1975). Evaluation through medical records and related information systems. In E. Struening & M. Guttentag (Eds.), Handbook of evaluation research Vol. 1 (pp. 397-484). Beverly Hills, CA: Sage Publications, Inc.

Weiss, C. (1975). Interviewing in evaluation research. In E. Struening & M. Guttentag (Eds.), Handbook of evaluation research Vol. 1 (pp. 355-396). Beverly Hills, CA: Sage Publications, Inc.

Weiss, E. (Ed.) (1972). Evaluating action programs: Readings in social action and education. Boston: Allyn & Bacon, Inc.

Yates, B. T. (1978). Improving the cost-effectiveness of obesity programs: Three basic strategies for reducing the cost per pound. International Journal of Obesity, 2, 249-266.

RECENT GUIDEBOOKS IN THIS SERIES

No.	Title
1	Microcomputers and Evaluation
2	Cost-Outcome Analysis: Measuring Costs
3	Microcomputers: Word Processing
4	Cost-Outcome Analysis: Measuring Outcomes
5	Microcomputers: Statistical Analysis Software
6	Investigative Journalism Techniques
7	Microcomputers: Data Base Management Software
8	Committee Hearings: Their Use in Evaluation
9	Microcomputers: Spreadsheet Software
10	Methods of Product Evaluation

These materials are in the public domain and may be reproduced without permission. The following acknowledgment is requested on materials which are reproduced: Developed by the Northwest Regional Educational Laboratory, Portland, Oregon.

Printed by the Northwest Regional Educational Laboratory, a private nonprofit corporation. The work upon which this publication is based was performed pursuant to Contract No. 400-80-0105 of the National Institute of Education. It does not, however, necessarily reflect the views of that agency.

Research on Evaluation Program
Northwest Regional Educational Laboratory
300 S. W. Sixth Avenue
Portland, OR 97204